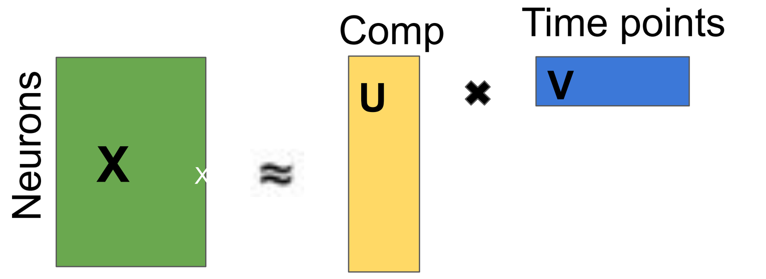
Ensemble Pursuit: an algorithm for finding overlapping clusters of correlated neurons in large-scale recordings

Maria Kesa, Carsen Stringer*, Marius Pachitariu*

Goal:

In order to inspire, confirm or reject theories about the brain we need to extract meaningful information from neural data. To find ensembles of co-activating cells we use a matrix factorization framework.



Algorithm:

Minimize:

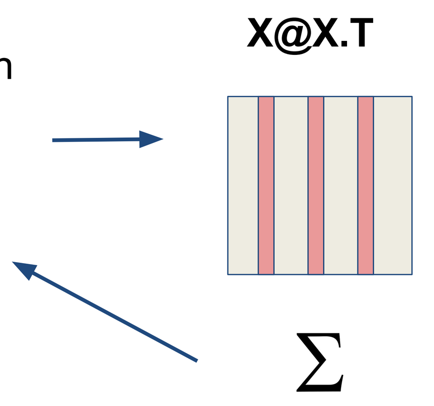
$$\text{Cost} = \|X - U\cdot V\|^2 + \boxed{\lambda\|U\|_0}$$

$$\Delta_j = \frac{\max(0, \vec{v}_j^T \vec{x})^2}{\|\vec{v}\|^2} - \lambda$$

For number of ensembles to fit:

Fit one ensemble:

New neuron index j, add to ensemble. Stop when cost delta is below or equal to 0.



Results:

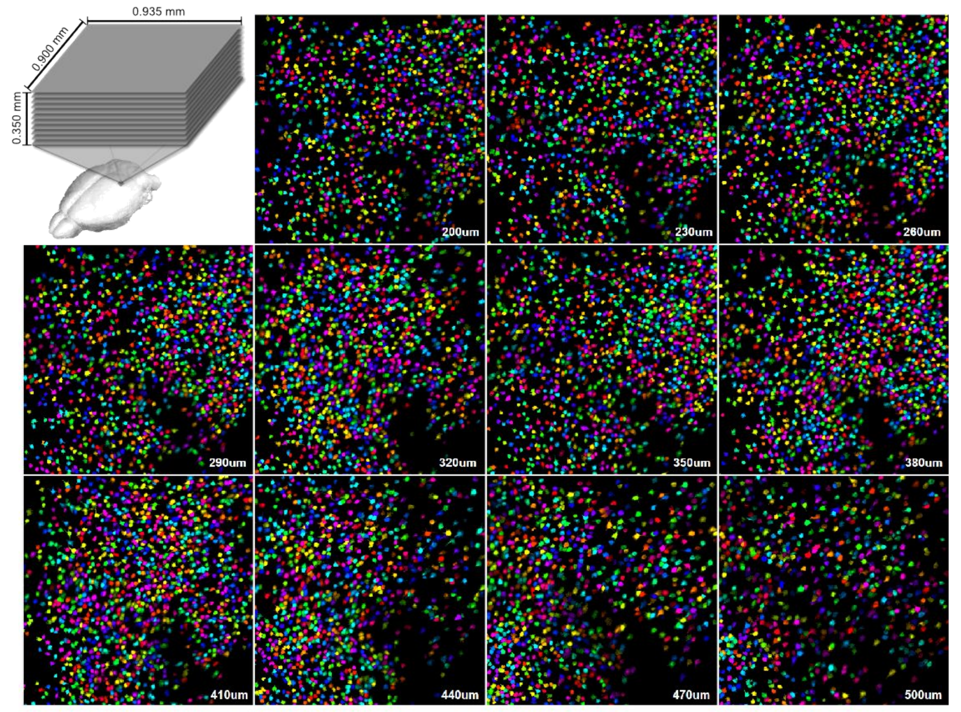
- Ensemble Pursuit is fast and accurate at finding ensembles in high dimensional neural data
- Stimulus information can be decoded from ensembles
- Ensembles of cells exhibit Gabor-like linear receptive fields

	Full data	Ensemble Pursuit	SparsePCA	PCA	NMF	LDA
knn accuracy (150)	0.37, se. 0.02	0.26, se. 0.04	0.36, se. 0.06	0.30, se. 0.05	0.21, se. 0.03	0.12, se. 0.04
Runtime (150)	-	7 min	124 min	2 sec	3 min	90 min
Median Sparsity, % of non-zeros	-	1%	60%	100%	71%	100%

Conclusions:

Ensemble Pursuit is an algorithm that finds ensembles in large-scale recordings and can thus be used as a stepping stone for further theoretical work.

Imaging 10,000 neurons simultaneously

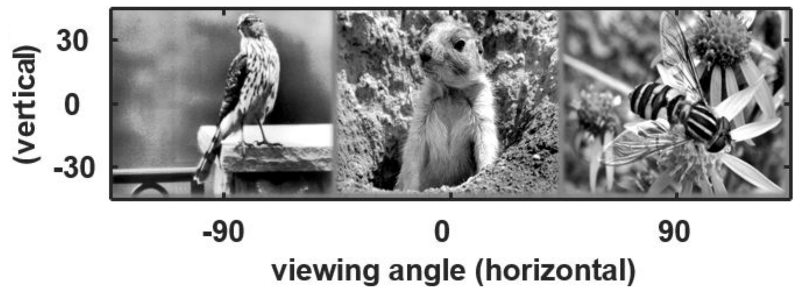


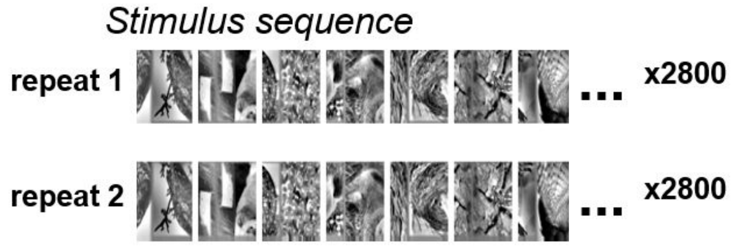
Recordings in mouse visual cortex of GCaMP6s (7 recordings, 6 mice, ~11000 neurons per recording)

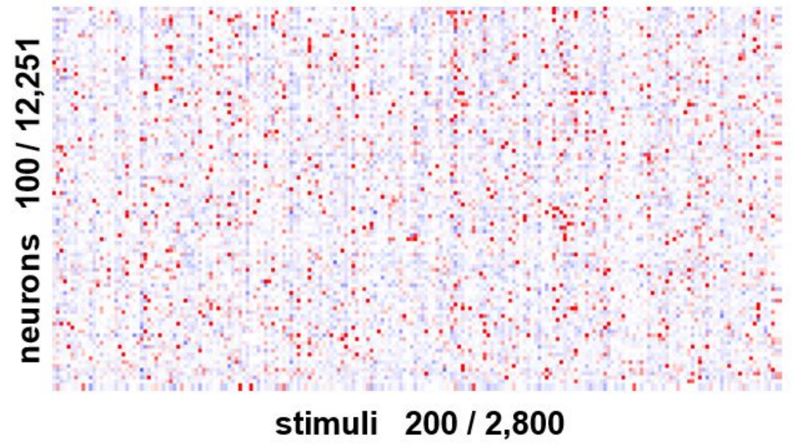
Data processed using **Suite2P** (Pachitariu et al, 2016, BiorXiv)

github.com/cortex-lab/Suite2P

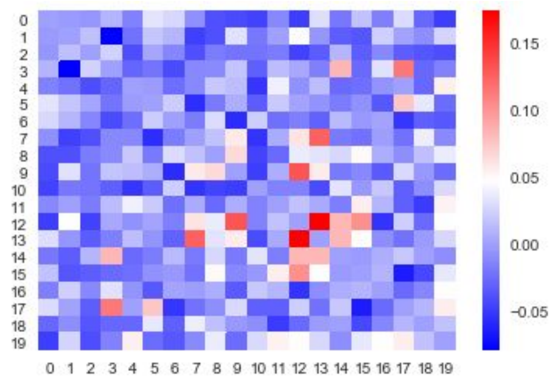
Responses to 2,800 natural images

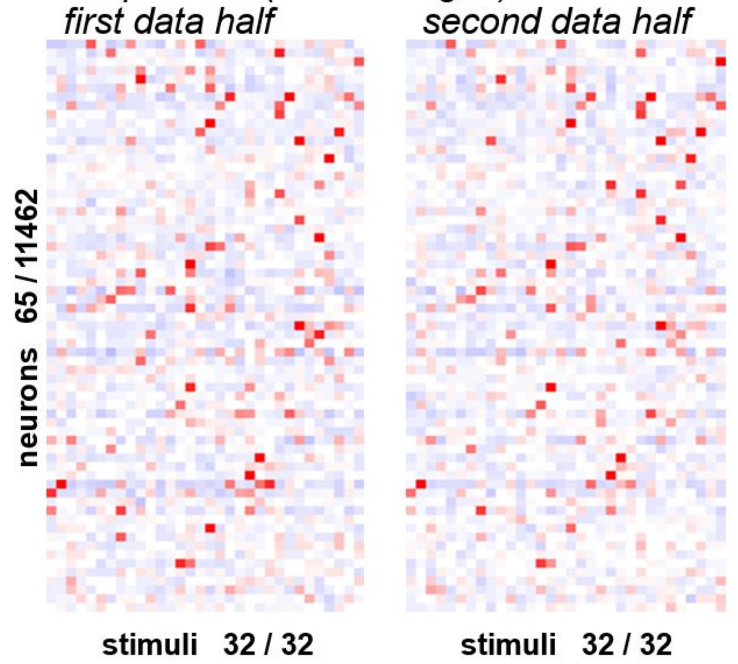




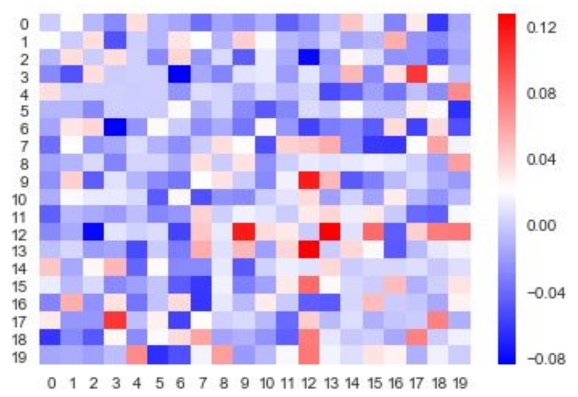


Covariance between 20 random neurons for first stimulus presentation





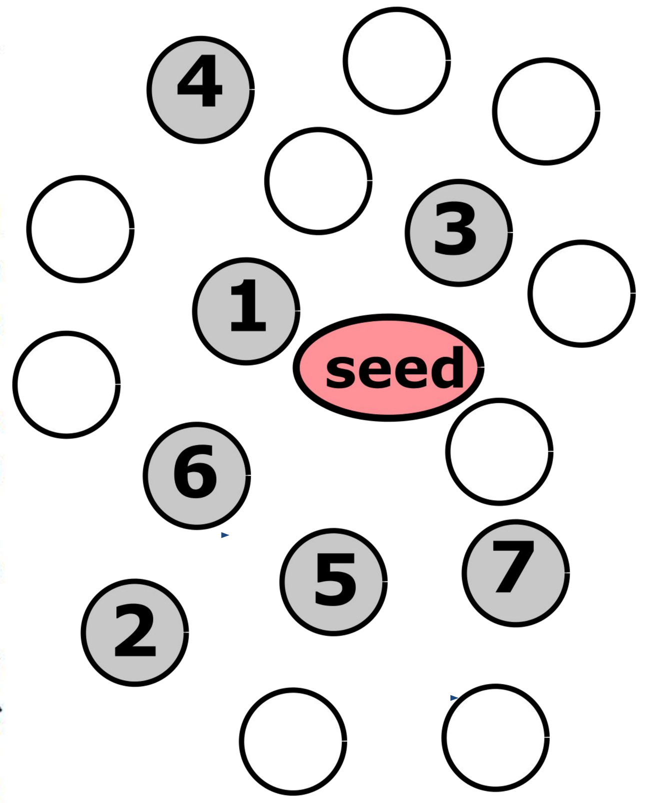
Covariance between the same 20 random neurons for second stimulus presentation

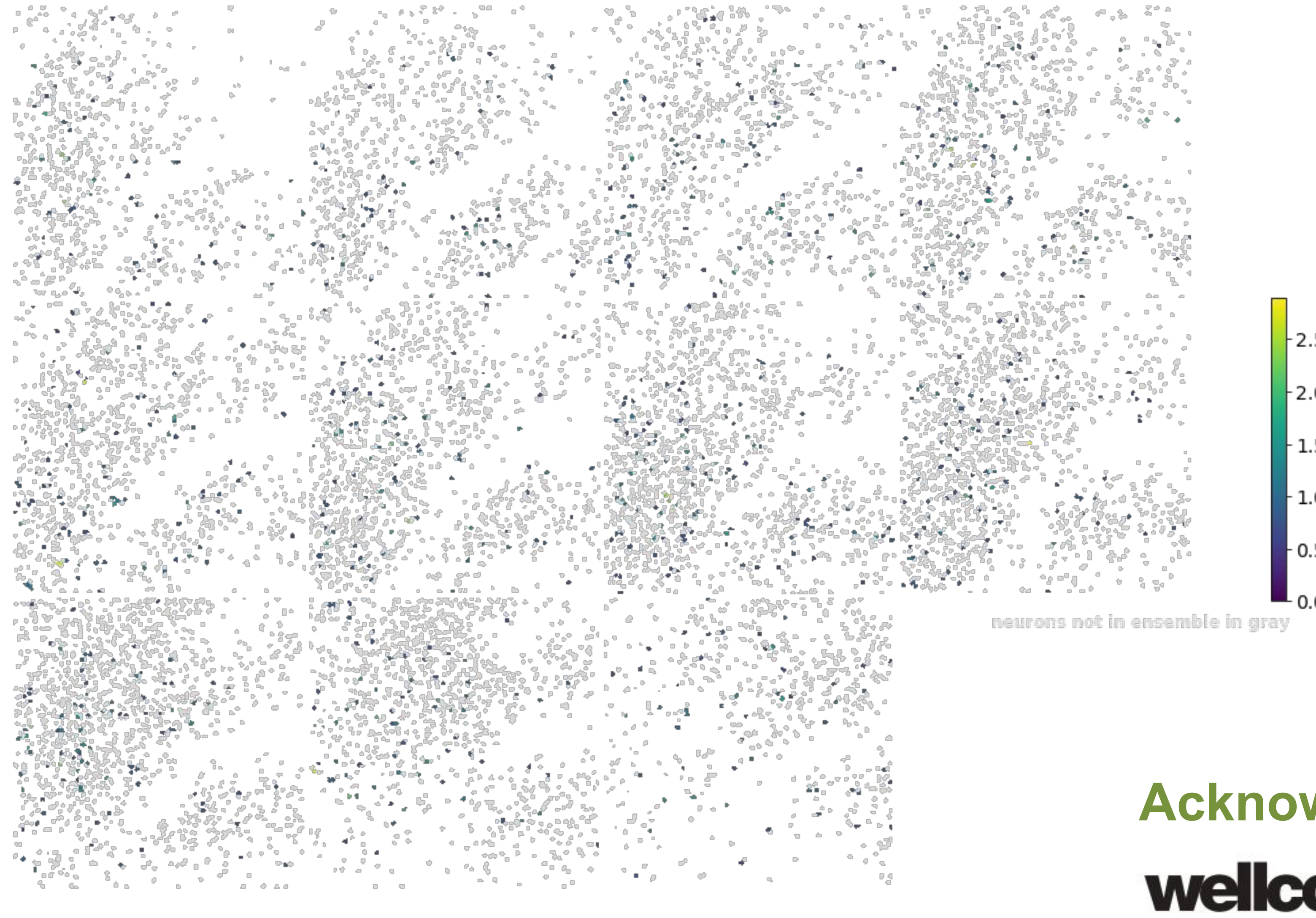


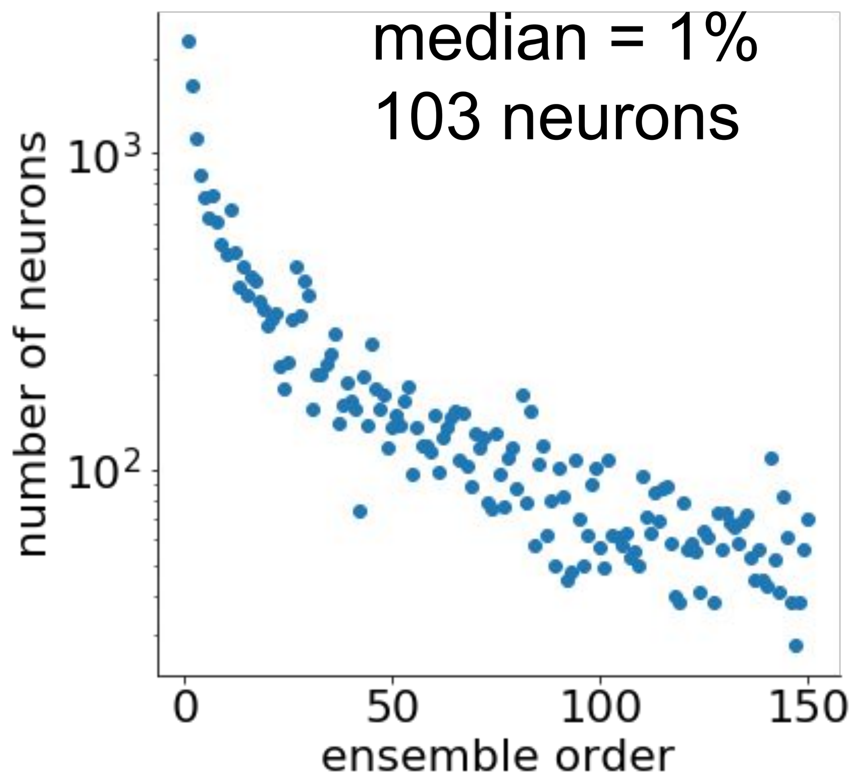
Finding overlapping clusters of neurons with Ensemble Pursuit

Algorithm 1 Ensemble pursuit algorithm

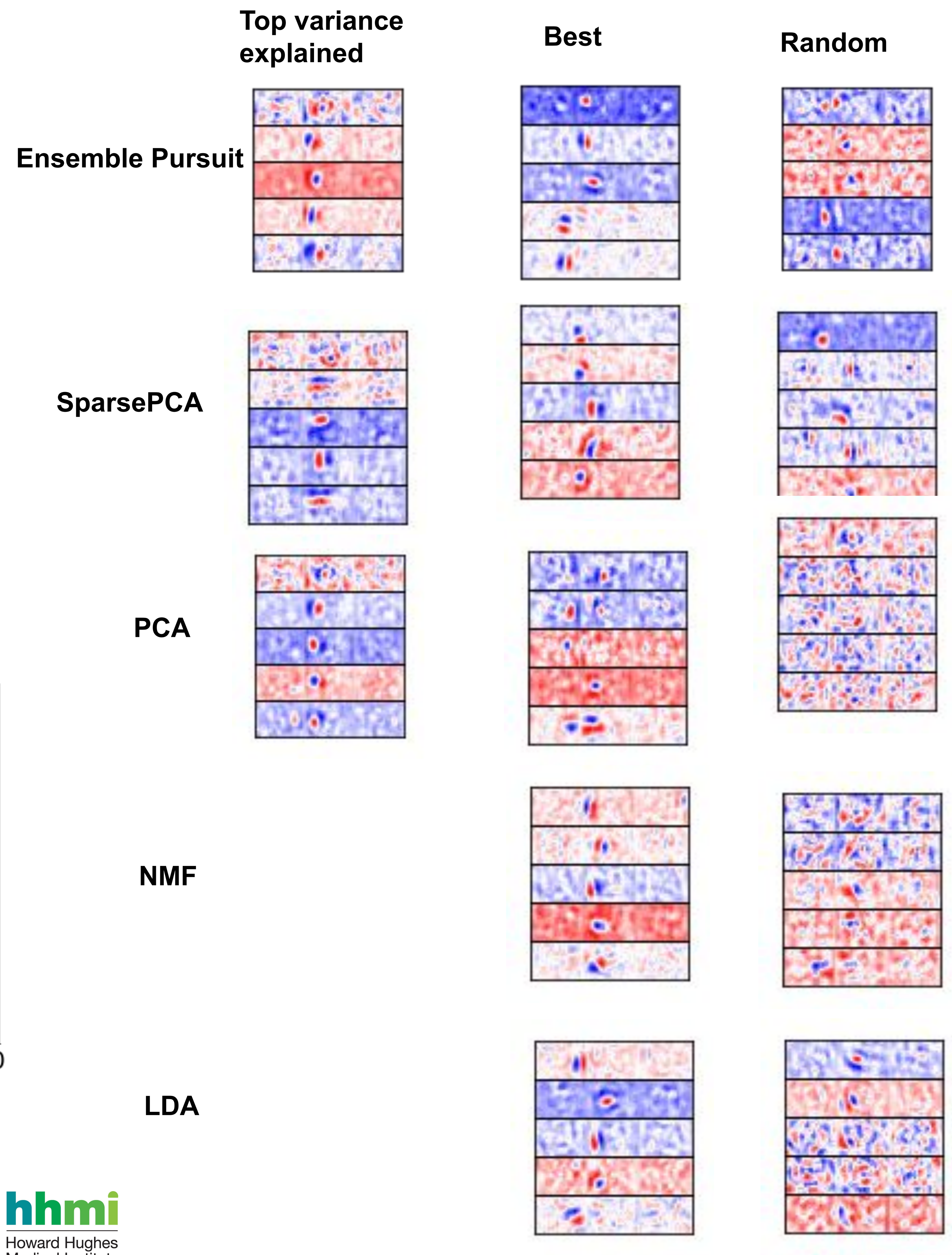
procedure FITENSEMBLES(X)
  $X \leftarrow$ matrix of neurons x timepoints
  $x_j \leftarrow$ row j of X
  $X = \text{zscore}(X)$ ▷ z-score X s.t. each neuron's activity has mean 0 and stddev 1
  for $n = 1 \rightarrow N_{ensembles}$ **do** ▷ extract $N_{ensembles}$ sequentially
    $j \leftarrow$ seed ▷ choose seed as neuron with highly correlated neighbors
    $\Delta_j = 1$
    $\vec{u} = [0...0]$ ▷ initialize vector $\vec{u}$ of length neurons with 0's
    $\vec{v} = \vec{x}_j$ ▷ initialize $\vec{v}$ as activity of seed neuron
    while $\Delta_j > 0$ **do**
      $w \leftarrow$ add j ▷ add new neuron to ensemble
      $n \leftarrow$ length of w
      $\vec{v} \leftarrow \sum_{k\in w} \vec{x}_k / n$ ▷ $\vec{v}$ is average of ensemble activity
      $j \leftarrow \max\{k \notin w \mid \max(0, \vec{x}_k^T \vec{v})^2\}$ ▷ choose new neuron most correlated to $\vec{v}$
      $\Delta_j \leftarrow \frac{\max(0, \vec{x}_j^T \vec{v})^2}{\|\vec{v}\|^2} - \lambda$ ▷ add neuron if cost decreases more than λ
    for $k \in w$ **do**
      $u_k \leftarrow \frac{\vec{x}_k^T \vec{v}}{\|\vec{v}\|^2}$ ▷ regress to find weight of neuron in ensemble
      $\vec{x}_k \leftarrow \vec{x}_k - u_k \vec{v}$ ▷ subtract ensemble contribution
    $U \leftarrow$ add $\vec{u}$
    $V \leftarrow$ add $\vec{v}$







Receptive fields of ensembles (ridge regression)



Acknowledgments

wellcome trust, SIMONS FOUNDATION, GATSBY, hhmi Howard Hughes Medical Institute